Causal Inference with an Interval Censored Exposure

by

Xinye Li

A thesis submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Master of Science

Baltimore, Maryland
April, 2019

Abstract

The purpose of this thesis is to use data collected on patients with severe pelvic fractures who received circumferential pelvic compression (CPC) to draw inference about the causal effect of the CPC placement time (since injury) on blood products use. Our approach focused on estimation of the probability of blood product use had all injured patients received CPC at time z after injury. A unique methodological challenge that is addressed is the interval-censoring of time of CPC placement. Our ultimate analysis, which is hampered by limited sample size and information on key confounders, did not find evidence that earlier binder placement reduced the risk of blood product use.

Thesis Committee

Primary Readers

Daniel Scharfstein (Primary Advisor)
Professor
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Renan Castillo
Associate Professor
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health

Acknowledgments

I would like to thank my advisor, Dr. Daniel Scharfstein, without whose unwavering commitment, generous support and help this thesis would not have been possible. Dan has taught me many important lessons, not only in statistics, but also with regards to rigorous scientific attitude towards research. I truly appreciate his patience with my frequent requests for help.

I would also like to thank the METRC Pilot-Bind team for sharing data and allowing me to use it as motivation for the methods developed in this thesis. I am also appreciative to my second reader, Dr. Renan Castillo, for providing insightful feedback on my thesis.

I also heartily thank peers in my academic cohort and my colleagues from the Johns Hopkins Biostatistics department for their advice and support.

Lastly, as always, I would like to thank my wonderful friends and family for their love and support over the last two years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Pelvic fractures in non-elderly adults are usually the result of high-impact trauma (Goins et al., 1992). Such fractures can often be accompanied by life-threatening internal bleeding (Rothenberger, Fischer, and Perry, 1978). One treatment that is considered safe, efficacious and simple to implement is called circumferential pelvic compression (CPC). CPC involves the application of radial inward forces through constriction of a binder or sheet to stabilize the pelvis. It reduces pelvic volume and impedes bleeding by facilitating of clot formation (White, Hsu, and Holcomb, 2009). CPC has been advocated as a first-line treatment for the severe pelvic injuries from multiple studies (Scott et al., 2013; see also Black et al., 2016; Bottlang et al., 2002; Hak, Smith, and Suzuki, 2009; Prasarn et al., 2013). While application of CPC often occurs in the hospital, some emergency medical services have trained personnel to apply it in the field.

The Pilot-BIND study retrospectively collected information on adults 18-64 years admitted to one of 23 trauma centers between August 1, 2015 and July 31, 2016, who (1) sustained a severe blunt or blast traumatic pelvic injury, (2)

had a severe pelvic fracture associated with internal bleeding, and (3) arrived alive at a hospital of definitive care within 12 hours of injury. For each patient, (a) use and timing (if applicable) of CPC, (b) use of blood products, and (c) severity of injury was to be recorded. The goal of this thesis is to use data from patients who received CPC to learn about the causal effect of the time (from injury) of CPC placement on the use of blood products. In addressing this goal, a key methodological difficulty is that time of CPC placement is not exactly known (i.e., coarsened) for some patients. Among the 240 patients with CPC placement (and recorded injury severity), time of placement was known exactly, interval censored and right censored for 121 (50.4%), 41 (17.1%), 78 (32.5%) patients, respectively. To address this problem, we build a casual inference procedure using the likelihood approach of Gomez, Espinal, and Lagakos (2003), who developed a method for estimating regression parameters in a generalized linear model with coarsened regressors.

This thesis is organized as follows. Chapter 2 introduces our methodology. Chapter 3 applies the methodology to the Pilot-BIND study. The last chapter is devoted to a discussion and directions for future research.

Chapter 2

Methods

2.1 Data Structure and Notation

We assume that the time scale for CPC placement is measured in minutes from injury. Let $Y(z)$ denote an individual's blood use (yes/no) if the CPC was placed at time z. Let Z be the time of CPC placement, assumed to lie in the set $S = \{s_1, \ldots, s_m\}$ with $m < \infty$. Let X be a measure of injury severity. Let L and R be the observed lower and upper bounds for Z that arise through some coarsening process. Let Y be the observed blood use, which is equal to $Y(Z)$. When $L = R$, Z is observed exactly; when L and R are both finite then Z is interval censored; and when L is finite and $R = \infty$, then Z is right-censored. The observable data for an individual is $O = (X, L, R, Y)$. We assume we observed n independent and identically distributed copies of O. When necessary, we will use the subscript i to denote the data for the ith individual.

2.2 Estimand, Assumptions and Modeling

The goal is to draw inference about $P[Y(z) = 1]$ as a function of z. This function will be estimated using the following assumptions:

Assumption 1: Z is independent of $Y(z)$ given $X = x$

This assumption says that, given information on injury severity, the time of CPC placement is randomized.

Under Assumption 1,

$$\begin{aligned} P[Y(z) = 1] &= \int P[Y(z) = 1|X = x]dF(x) \\ &\overset{A1}{=} \int P[Y(z) = 1|Z = z, X = x]dF(x) \\ &= \int P[Y = 1|Z = z, X = x]dF(x) \end{aligned}$$

where $F(\cdot)$ is the distribution of X. To estimate $P[Y(z) = 1]$, we need to estimate $P[Y = 1|Z = z, X = x]$. This will require two additional assumptions about the coarsening mechanism.

Assumption 2: $P[Y = 1|Z = z, L = l, R = r, X = x] = P[Y = 1|Z = z, X = x], l \leq z \leq r$

This assumption says that, given information on injury severity and the true CPC placement time, the coarsening process provides no information about the use of blood products.

Assumption 3: $P[Z = z|L = l, R = r, X = x] = \frac{P[Z=z|X=x]}{P[l\leq Z\leq r|X=x]}, l \leq z \leq r$

This assumption says that, given information on injury severity, the process that coarsens the CPC placement time provides no information about the true

placement time beyond knowing that the true time lies in the interval.

We also introduce two models to increase the efficiency of estimation.

Model 1: logit $P[Y = 1|Z = z, X = x] = \alpha_0 + \alpha_1 z + \alpha_2 x$

Model 2: logit $P[Z = z|Z \geq z, X = x] = \beta_0 + \beta_1 z + \beta_2 x$

Model 2 is the discrete-time proportional hazards discussed by Allison (1982). The induced model for $P[Z = z|X]$ can be derived, inductively, using the fact that

$$\frac{P[Z = z|X = x]}{1 - \sum_{z'<z} P[Z = z'|X = x]} = P[Z = z|Z \geq z, X = x]$$

Note that $P[Z = s_1|X = x] = P[Z = s_1|Z \geq s_1, X = x]$. Suppose that $P[Z = s_j|X = x]$ is known for all $j = 1, \ldots, k$. Then,

$$P[Z = s_{k+1}|X = x] = P[Z = s_{k+1}|Z \geq s_{k+1}, X = x] \left\{ 1 - \sum_{j=1}^{k} P[Z = s_j|X = x] \right\}$$

2.3 Likelihood of the Observed Data

Let $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ and $\beta = (\beta_0, \beta_1, \beta_2)$. Note that

$$
\begin{aligned}
&P[Y = y | L = l, R = r, X = x] \\
&= \sum_{l \le z \le r} P[Y = y, Z = z | L = l, R = r, X = x] \\
&= \sum_{l \le z \le r} P[Y = y | Z = z, L = l, R = r, X = x] P[Z = z | L = l, R = r, X = x] \\
&\overset{A2}{=} \sum_{l \le z \le r} P[Y = y | Z = z, X = x] P[Z = z | L = l, R = r, X = x] \\
&\overset{A3}{=} \sum_{l \le z \le r} P[Y = y | Z = z, X = x] \frac{P[Z = z | X = x]}{P[l \le Z \le r | X = x]}
\end{aligned}
$$

and

$$P[L = l, R = r | X = x] = P[l \le Z \le r | X = x] P[L = l, R = r | l \le Z \le r, X = x].$$

Thus,

$$
\begin{aligned}
&P[Y = y, L = l, R = r | X = x] \\
&= \left\{ \sum_{l \le z \le r} P[Y = y | Z = z, X = x] P[Z = z | X = x] \right\} \times \\
&\qquad \underbrace{P[L = l, R = r | l \le Z \le r, X = x]}_{\text{Ancillary}}.
\end{aligned}
$$

where, under Assumption (3), the latter probability does not contain information about α and β (see Gill, van der Laan, and Robins, 1997).

With these results, the key components of the observed data likelihood for person i can be written as:

$$\mathcal{L}_i(\alpha,\beta) = \sum_{L_i \le z \le R_i} P[Y = Y_i | Z = z, X = X_i; \alpha] P[Z = z | X = X_i; \beta]$$

The associated log-likelihood and scores are

$$\ell_i(\alpha,\beta) = \log\left\{ \sum_{L_i \le z \le R_i} P[Y = Y_i | Z = z, X = X_i; \alpha] P[Z = z | X = X_i; \beta] \right\}$$

$$S_{\alpha,i}(\alpha,\beta) = \frac{\sum_{L_i \le z \le R_i} \frac{\partial P[Y=Y_i|Z=z,X=X_i;\alpha]}{\partial \alpha} P[Z = z | X = X_i; \beta]}{\sum_{L_i \le z \le R_i} P[Y = Y_i | Z = z, X = X_i; \alpha] P[Z = z | X = X_i; \beta]}$$

$$S_{\beta,i}(\alpha,\beta) = \frac{\sum_{L_i \le z \le R_i} P[Y = Y_i | Z = z, X = X_i; \alpha] \frac{\partial P[Z=z|X=X_i;\beta]}{\partial \beta}}{\sum_{L_i \le z \le R_i} P[Y = Y_i | Z = z, X = X_i; \alpha] P[Z = z | X = X_i; \beta]}$$

Let $S_i(\alpha,\beta) = (S_{\alpha,i}(\alpha,\beta)', S_{\beta,i}(\alpha,\beta)')'$. The likelihood for the observed data for all persons is $\mathcal{L}(\alpha,\beta) = \prod_{i=1}^n \mathcal{L}_i(\alpha,\beta)$. The associated log-likelihood is $\sum_{i=1}^n \ell_i(\alpha,\beta)$ and score vector is

$$S(\alpha,\beta) = \sum_{i=1}^n (S_{\alpha,i}(\alpha,\beta)', S_{\beta,i}(\alpha,\beta)')'$$

The observed information matrix is $I(\alpha,\beta) = \sum_{i=1}^n S_i(\alpha,\beta) S_i(\alpha,\beta)'$.

2.4 Estimation and Inference

Let $\widehat{\alpha}$ and $\widehat{\beta}$ be the maximum likelihood estimator for α and β, respectively. Using the theory of maximum likelihood, we know that $\widehat{\alpha}$ and $\widehat{\beta}$ solves $S(\alpha,\beta) = 0$ and

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} \approx N\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, I(\widehat{\alpha},\widehat{\beta})^{-1} \right)$$

Unfortunately, it is very complicated to compute the score vector and the information matrix. Thus, we maximize log-likelihood directly using the `optim` function with the `BFGS` algorithm in `R`. The BFGS algorithm is a quasi-Newton approach which iteratively optimizes a nonlinear function without computing the Hessian (Curtis and Que, 2015). We also compute an alternative version of the observed information matrix using the Hessian matrix of the negative log-likelihood evaluated at $(\widehat{\alpha}, \widehat{\beta})$. Denote this version as $J(\widehat{\alpha}, \widehat{\beta})$. The Hessian can be computed using numerical differentiation using the `numderiv` function in `R`.

Then, we estimate $P[Y(z) = 1]$ by

$$\widehat{P}[Y(z) = 1] = \int P[Y = 1 | Z = z, X = x; \widehat{\alpha}] d\widehat{F}(x)$$

where $\widehat{F}$ is the empirical distribution of X. This estimator can be expressed as

$$\widehat{P}[Y(z) = 1] = \frac{1}{n} \sum_{i=1}^{n} \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\} \tag{2.1}$$

Treating the distribution of X as the fixed true distribution of X, the large sample distribution of $\widehat{P}[Y(z) = 1]$ can be computed using the delta method. That is,

$$\widehat{P}[Y(z) = 1] \approx N(P[Y(z) = 1], g(\widehat{\alpha})' I(\widehat{\alpha}, \widehat{\beta})^{-1} g(\widehat{\alpha}))$$

where

$$g(\widehat{\alpha}) = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\} \{1 - \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\}\} \\ \frac{1}{n} \sum_{i=1}^{n} \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\} \{1 - \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\}\} z \\ \frac{1}{n} \sum_{i=1}^{n} \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\} \{1 - \text{expit} \{\widehat{\alpha}_0 + \widehat{\alpha}_1 z + \widehat{\alpha}_2 X_i\}\} X_i \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In operationalizing the construction of confidence intervals, we use $J(\widehat{\alpha}, \widehat{\beta})$ instead of $I(\widehat{\alpha}, \widehat{\beta})$. That is, a 95% confidence interval for $P[Y(z) = 1]$ takes the form:

$$[\widehat{P}[Y(z) = 1] - 1.96 g(\widehat{\alpha})' J(\widehat{\alpha}, \widehat{\beta})^{-1} g(\widehat{\alpha}), \widehat{P}[Y(z) = 1] + 1.96 g(\widehat{\alpha})' J(\widehat{\alpha}, \widehat{\beta})^{-1} g(\widehat{\alpha})].$$

Chapter 3

Data Analysis

3.1 Preliminaries

The Pilot-BIND study was designed to retrospectively collect information on the timing of key sentinel events, including time of injury, times of EMS arrival and departure, times of transfer hospital arrival and departure (if applicable), and time of definitive care hospital arrival. Figure 3.1 displays the timeline of events, with the solid dots representing the data collection elements. Unfortunately, time of event data was missing for many patients. As a pre-processing step (not discussed here), the missing time elements were multiply imputed to create five "complete" datasets.

In addition, the use and timing (if applicable) of CPC was recorded. Among many of the patients who received CPC, however, the time of CPC was missing. For these patients, the provider (i.e., EMS, transfer hospital, definitive care hospital) of CPC was known. Thus, for each imputed dataset, the time (from injury) of CPC will be interval censored for those whose source is EMS or transfer hospital and right censored for those whose source is the

definitive hospital. All binders were assumed to be placed no later than 1000 minutes after injury, so that right censored patients are considered to be interval censored between time (from injury) of definitive hospital arrival and 1000 minutes.

Our analysis focuses on 240 patients who received CPC and had information on injury severity. Injury severity is characterized by two variables: injury severity score (ISS) and indicator of an open fracture. ISS is defined as the sum of squares of the highest abbreviated injury scale (AIS) code from three of the most severely injured body regions. The six body regions are: (1) head/neck, (2) face, (3) chest, (4) abdominal organs/lumbar spine, (5) extremities/pelvic skeleton and (6) external. AIS is essentially a seven point scale, with 0 coded as no injury, 1 coded as minor, 2 as moderate, 3 as serious, 4 as severe, 5 as critical and 6 as maximal (i.e., untreatable). If any body region has an AIS of 6, then the ISS is set equal to 75. Therefore, the ISS score ranges from 1 to 75 (Baker et al., 1974).

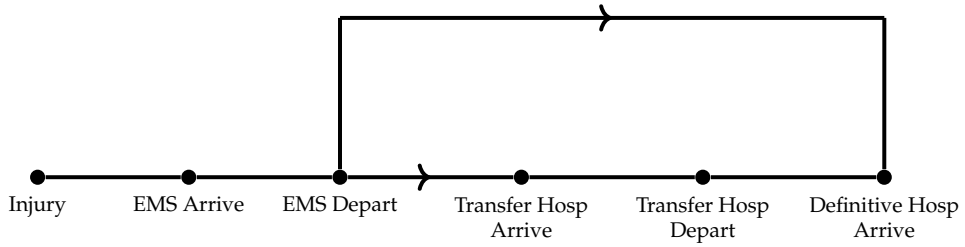


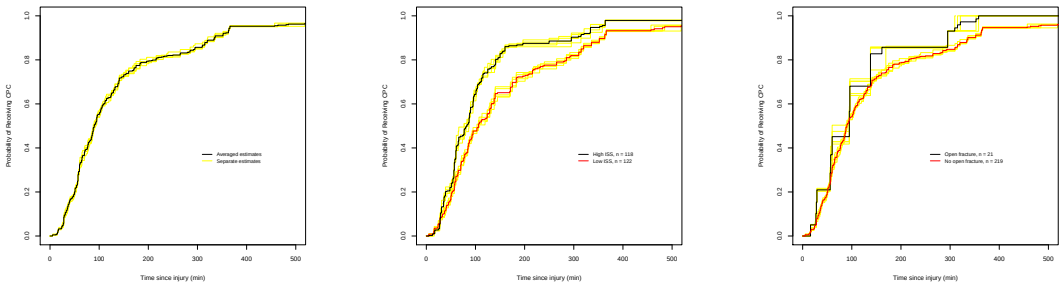
Figure 3.1: Timeline

3.2 Descriptive Analyses

Among the 240 patients, the average (median) ISS is 28.1 (25.0); the first and third quartiles are 17.8 and 35.0, respectively. The minimum and maximum scores are 4 and 66. In addition, 21 (8.75%) have open fractures. Table 3.1 displays summary statistics for observed CPC data from the five imputed datasets. Figure 3.2a displays the estimated (using the method of Turnbull, 1976) cumulative distribution function of time of CPC for each imputed dataset (yellow) and the averaged across datasets (black). The first, second (median) and third quartiles of the averaged distribution are 57, 91 and 160, respectively. Figure 3.2b displays the estimated cumulative distribution function of time of CPC, stratified by low (below median) and high (at or above median) ISS score. The figure clearly shows that more severely injured patients are more likely to get CPC earlier. Figure 3.2c displays the estimated cumulative distribution function of time of CPC, stratified by open fracture status (yes/no). Consistent with the results from Figure 3.2b, patients with open fractures are more likely to get CPC earlier, though the interpretation is limited due to the small number of patients with open fractures.

Table 3.1: CPC time/width (rounded to the nearest minute) from five imputed datasets

		min	mean(sd)	max
Z Exact Time (n=121)	Imputation 1	5	133 (150)	980
	Imputation 2	5	134 (148)	980
	Imputation 3	5	131 (150)	980
	Imputation 4	5	131 (148)	980
	Imputation 5	5	133 (150)	980
Z Interval-Censored Width (n = 41)	Imputation 1	0	91 (105)	445
	Imputation 2	0	90 (85)	393
	Imputation 3	1	80 (82)	305
	Imputation 4	1	101 (129)	647
	Imputation 5	0	84 (89)	404
Z Right-Censored Width (n = 78)	Imputation 1	579	934 (78)	996
	Imputation 2	579	939 (73)	997
	Imputation 3	579	933 (79)	996
	Imputation 4	579	933 (80)	997
	Imputation 5	579	933 (79)	996



(a) Overall **(b)** By ISS **(c)** By open fracture

Figure 3.2: Estimated probability of receiving CPC over time

3.3 Formal Analyses

In our analyses, we used minutes as the time scale, set $m = 1000$ and let $S = \{1, \ldots, 1000\}$. We fit **Model (1)** and **Model (2)** described in Chapter 2, with $X = (X_1, X_2)$, X_1 equal to ISS and X_2 equal to the indicator of open fracture . We also considered another version of these models with more flexible parameterizations for z. Specifically, we considered:

Model 1': logit $P[Y = 1|Z = z, X = X] = f(z; \alpha_1) + \alpha_2 x$

Model 2': logit $P[Z = z|Z \geq z, X = X] = f(z; \beta_1) + \beta_2 x$

where $\alpha_1 = (\alpha_{1,0}, \alpha_{1,1}, \alpha_{1,2}, \alpha_{1,3}, \alpha_{1,4})$, $\beta_1 = (\beta_{1,0}, \beta_{1,1}, \beta_{1,2}, \beta_{1,3}, \beta_{1,4})$,

$$f(z; \alpha_1) = \begin{cases} \alpha_{1,0}\frac{z-z_1}{z_0-z_1} + \alpha_{1,1}\frac{z-z_0}{z_1-z_0} & z \in [z_0, z_1] \\ \alpha_{1,1}\frac{z-z_2}{z_1-z_2} + \alpha_{1,2}\frac{z-z_1}{z_2-z_1} & z \in [z_1, z_2] \\ \alpha_{1,2}\frac{z-z_3}{z_2-z_3} + \alpha_{1,3}\frac{z-z_2}{z_3-z_2} & z \in [z_2, z_3] \\ \alpha_{1,3}\frac{z-z_4}{z_3-z_4} + \alpha_{1,4}\frac{z-z_3}{z_4-z_3} & z \in [z_3, z_4], \end{cases}$$

$f(z; \beta_1)$ is the same as $f(z; \alpha_1)$ with the exception that the α_1 vector is replaced by the β_1 vector, $z_0 = 0$, $z_1 = 57$, $z_2 = 91$, $z_3 = 160$, $z_4 = 1000$. The functions $f(z; \beta_1)$ and $f(z; \alpha_1)$ are called piecewise linear splines (Qiu, 2013). The parameters $\alpha_{1,0}$ and $\beta_{1,0}$ are intercepts.

The estimation approach described in Chapter 2 was applied to each imputed dataset. The estimated regression parameters from the five imputed datasets were averaged to obtain an overall estimate. The variance-covariance matrix of the overall estimator was estimated as the sum of (1) the average of the estimated variance-covariance matrices across the five imputed datasets and (2) $(1 + 1/5)$ times the sample variance-covariance of the estimated regression parameters from the five imputed datasets (Rubin, 1996). The estimator

of $P[Y(z) = 1]$ was obtained by using Equation 2.1 with the overall estimator of α; the variance of this estimator was estimated using the delta method with the estimated variance-covariance matrix of the overall estimator of the regression coefficients. This approach was modified appropriately to handle to the piecewise-linear models described above. When fitting the regression models, ISS was divided by 66 and Z was divided by 1000; this scaling was employed for computational convenience.

The estimated regression coefficients along with standard errors and 95% confidence intervals from Models 1 and 2 are displayed in Table 3.2. Model 1 indicates associations between ISS and open fracture with the need for blood products. A one point (unit) increase in ISS was associated with an increase in odds of receiving blood products of 6% (odds ratio: 1.06, 95% CI: [1.03, 1.08]). The odds of receiving blood product among patients with open fractures was estimated to be 5.1 (95% CI: [1.1, 23.8]) times the odds among patients who did not have an open fracture. While the sign of the relationship between blood product use and time of CPC placement was positive, it was small and coupled with large uncertainty. Figure 3.3 presents the estimated $P[Y(z) = 1]$ as a function of z and Figure 3.4 (a) incorporates the 95% point-wise confidence interval. The estimated function is essentially flat at approximately 64% and the 95% pointwise confidence intervals are wide.

Table 3.2: Regression Coefficient Estimates (Standard Errors, 95% Confidence Intervals) for Models 1 and 2.

	Model 1				Model 2		
	Estimate	S.E.	95% CI		Estimate	S.E.	95% CI
α_0	-1.08	0.41	(-1.89, -0.27)	β_0	-5.20	0.18	(-5.56, -4.84)
α_1	0.02	1.15	(-2.23, 2.28)	β_1	0.06	0.47	(-0.87, 0.98)
$\alpha_{2,1}$	3.92	0.86	(2.24, 5.61)	$\beta_{2,1}$	0.38	0.37	(-0.33, 1.10)
$\alpha_{2,2}$	1.60	0.78	(0.08, 3.13)	$\beta_{2,2}$	0.21	0.28	(-0.33, 0.76)

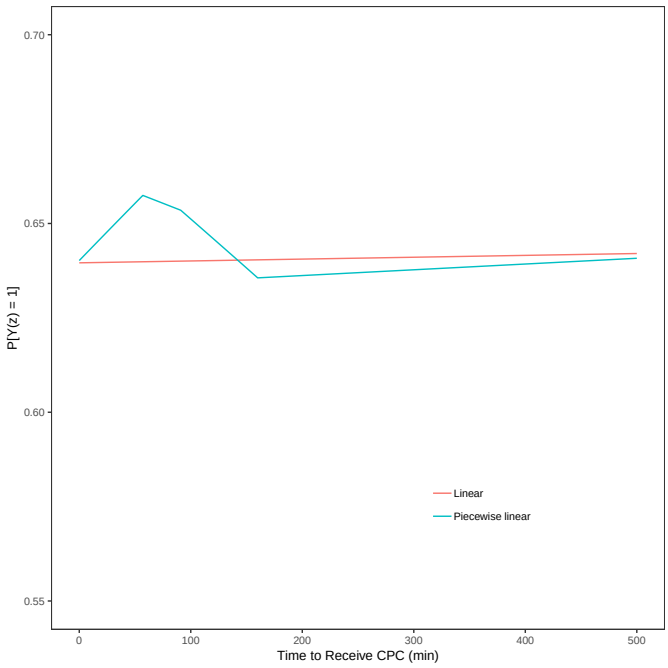
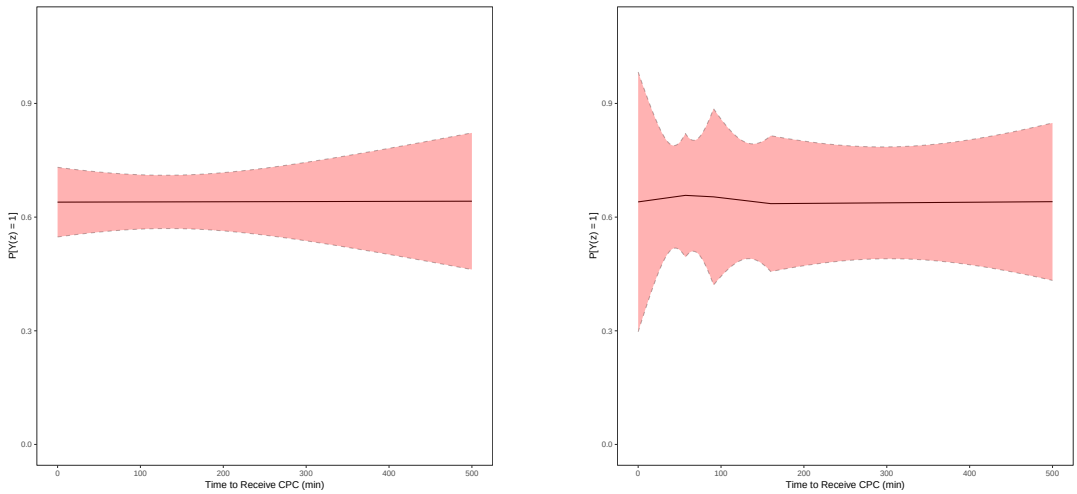


Figure 3.3: Estimated P[Y(z) = 1] from Models 1/2 (red) and Models 1′/2′ (blue)



(a) Model 1/2 **(b)** Model 1′/2′

Figure 3.4: Estimates with 95% pointwise confidence interval

The estimated regression coefficients along with standard errors and 95% confidence intervals from Models 1′ and 2′ are displayed in Table 3.3. Model 1′ indicates similar associations between ISS and open fracture with the need for blood products as in Model 1. Figure 3.3 presents the estimated $P[Y(z) = 1]$ as a function of z and Figure 3.4 (b) incorporates the 95% point-wise confidence interval. The estimated function has an interesting shape: the probability increases during the first 57 minutes after injury, then decreases during the interval $(57 \text{ minutes}, 160 \text{ minutes}]$ and then flattens out after 160 minutes. The range, however, of probability changes is quite small and the uncertainty is very large. Thus, the results are consistent with no effect on time of CPC placement on the need for blood products.

Table 3.3: Regression Coefficient Estimates (Standard Errors, 95% Confidence Intervals) for Models 1′ and 2′.

	Model 1′				Model 2′		
	Estimate	S.E.	95% CI		Estimate	S.E	95% CI
$\alpha_{1,0}$	-0.99	0.92	(-2.79, 0.81)	$\beta_{1,0}$	-6.80	0.39	(-7.57, -6.04)
$\alpha_{1,1}$	-0.90	0.48	(-1.84 0.04)	$\beta_{1,1}$	-4.24	0.24	(-4.72, -3.76)
$\alpha_{1,2}$	-0.92	0.64	(-2.17, 0.34)	$\beta_{1,2}$	-4.53	0.31	(-5.14, -3.93)
$\alpha_{1,3}$	-1.01	0.50	(-2.00, -0.02)	$\beta_{1,3}$	-5.21	0.26	(-5.72, -4.70)
$\alpha_{1,4}$	-0.95	1.35	(-3.59, 1.70)	$\beta_{1,4}$	-5.29	0.59	(-6.44, -4.13)
$\alpha_{2,1}$	3.68	0.83	(2.05, 5.30)	$\beta_{2,1}$	0.17	0.38	(-0.57, 0.91)
$\alpha_{2,2}$	1.63	0.78	(0.10, 3.17)	$\beta_{2,2}$	0.22	0.28	(-0.32, 0.77)

Chapter 4

Discussion and Conclusion

In this thesis, we sought to illuminate the causal effect of timing of binder placement on the need for blood products. Our approach focused on estimation of the probability of blood product use, had all injured patients received CPC at time z, after injury. A unique aspect of our approach was handling of an interval-censored exposure (i.e., time of CPC placement). Our analysis did not find evidence that earlier binder placement reduced the risk of blood product use.

A key limitation of this analysis was the lack of information on variables that might plausibly confound the relationship between time of CPC placement and use of blood products. We were only able to adjust for non-specific injury severity variables through ISS and an indicator of open fracture. It would have been more desirable to account for pelvis-specific injury variables that are associated with bleeding, but these were not reliably collected.

Another key limitation is that our targeted quantity of inference does not refer to an actionable treatment policy. That is, it is not feasible to place a binder at exactly time z for all patients. Rather, our analysis attempts to

understand the causal effect of binder placement at a more "basic science" level.

An alternative approach would have been to estimate the causal effect of a treatment policies such as (1) CPC placement prior to EMS departure or (2) CPC placement within z^* minutes of injury. Evaluating such effects are not straightforward for at least two key reasons. First, due to interval censoring, we do not know, for all patients, whether they were observed to adhere to the policy or not. Second, there may be violations of stable unit treatment value assumption (SUTVA), which underlies classic causal inference techniques (Rubin, 1980). SUTVA means that there should not be "multiple versions of treatment" (Rubin, 1986). For these policies, however, there can be variation in the time of CPC placement within an individual that can, in theory, impact the need for blood products. To address this issue, it might be useful to consider "stochastic interventions", where time of CPC placement is drawn according to some fixed distribution (Munoz and van der Laan, 2011).

In summary, we have developed and applied a causal inference methodology where the exposure is interval censored. Our ultimate inferences, when applied to the Pilot-BIND study, were highly uncertain and did not allow us to draw definite conclusions about the effect of time of CPC placement on the need for blood products. Nonetheless, we see interesting directions for methodological research to address the treatment policy questions discussed above.

References

Goins, WA, Rodriguez, A, Lewis, J, Brathwaite, CE, and James, E (1992). "Retroperitoneal hematoma after blunt trauma". In: *Surgery, Gynecology & Obstetrics* 174.4, pp. 281–290.

Rothenberger, DA, Fischer, RP, and Perry, Jr. JF (1978). "Major vascular injuries secondary to pelvic fractures: an unsolved clinical problem". In: *American Journal of Surgery* 136.6, pp. 660–662.

White, CE, Hsu, JR, and Holcomb, JB (2009). "Haemodynamically unstable pelvic fractures". In: *Injury* 40.10, pp. 1023–1030.

Scott, I, Porter, K, Laird, C, Greaves, I, and Bloch, M (2013). "The prehospital management of pelvic fractures: Initial consensus statement". In: *Emergency Medicine Journal* 30.12, pp. 1770–1772.

Black, SR, Sathy, AK, Jo, C, Wiley, MR, Minei, JP, and Starr, AJ (2016). "13 year experience at a single trauma center using a multidisciplinary institutional protocol". In: *Journal of Orthopaedic Trauma* 30.1, pp. 22–28.

Bottlang, M, Simpson, T, Sigg, J, Krieg, JC, Madey, SM, and Long, WB (2002). "Noninvasive reduction of open-book pelvic fractures by circumferential compression". In: *Journal of Orthopaedic Trauma* 16.6, pp. 367–373.

Hak, DJ, Smith, WR, and Suzuki, T (2009). "Management of hemorrhage in life-threatening pelvic fracture". In: *Journal of the American Academy of Orthopaedic Surgeons* 17.7, pp. 447–457.

Prasarn, ML, Conrad, B, Small, J, Horodyski, M, and Rechtine, GR (2013). "Comparison of circumferential pelvic sheeting versus the T-POD on unstable pelvic injuries: A cadaveric study of stability". In: *Injury* 44.12, pp. 1756–1759.

Gomez, G, Espinal, A, and Lagakos, SW (2003). "Inference for a linear regression model with an interval-censored covariate". In: *Statistics in Medicine* 22.1, pp. 409–425.

Allison, P (1982). "Discrete-time methods for the analysis of event histories". In: *Sociological Methodology* 13.3, pp. 61–98.

Gill, RD, van der Laan, MJ, and Robins, JM (1997). "Coarsening at random: Characterizations, conjectures, counter-examples". In: *Proceedings of the First Seattle Symposium in Biostatistics*. Springer, pp. 255–294.

Curtis, FE and Que, X (2015). "A quasi-Newton algorithm for nonconvex, nonsmooth optimization with global convergence guarantees". In: *Mathematical Programming Computation* 7.4, pp. 399–428.

Baker, SP, O'Neill, B, Haddon, W, and Long, WB (1974). "The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care". In: *Journal of Trauma - Injury, Infection and Critical Care* 14, pp. 187–196.

Turnbull, BW (1976). "The empirical distribution function with arbitrarily grouped censored and truncated data". In: *Journal of the Royal Statistical Society*. Series B (Methodological) 38, pp. 290–295.

Qiu, J (2013). *Interpolation by spline*. https://www.math.uh.edu/~jingqiu/math4364/spline.pdf. (Visited on 04/10/0019).

Rubin, DB (1996). "Multiple imputation after 18+ years". In: *Journal of the American Statistical Association* 91.434, pp. 476–477.

Rubin, DB (1980). "Comment on: 'Randomization analysis of experimental data in the fisher randomization test' by D. Basu." In: *Journal of the American Statistical Association*.

Rubin, DB (1986). "Which ifs have causal answers? Comment on: 'Statistics and causal inference' by P. Holland." In: *Journal of the American Statistical Association*.

Munoz, ID and van der Laan, MJ (2011). "Population intervention causal effects based on stochastic interventions". In: *Biometrics* 68.2, pp. 476–477.

Xinye Li

217-418-9952/ xli175@jhmi.edu/ https://www.linkedin.com/in/lixinye

EDUCATION

Johns Hopkins Bloomberg School of Public Health Expected: May 2019

- Sc.M. in Biostatistics Cumulative GPA: 3.76/4.0
  Advisor: Daniel O. Scharfstein, Ph.D.
- Thesis: Causal Inference with an Interval-censored Exposure
- Coursework: Machine Learning, Causal Inference, Longitudinal Data, and Bayesian Statistics

University of Illinois at Urbana-Champaign May 2017

- B.S. with honor in Applied Mathematics & Statistics, Cum Laude Cumulative GPA: 3.82/4.0
- Coursework: Survival Analysis, Time Series Analysis, and Epidemiology

EXPERIENCE

Major Extremity Trauma Research Consortium (METRC) **Baltimore, MD**
***Data Analyst**; Supervisor: Craig Remenapp, Senior Study Manager* Sep 2018-

- Extract reviews of trauma patients from the REDCap and conduct data management in R
- Program the logic of determining consensus on the reviews of each case from physicians
- Develop R-shiny Apps to interactively present tablet and figure outputs for each patient

IPSEN Bioscience Inc. **Cambridge, MA**
***Biometry Intern**; Supervisor: Tiffany Wang, Principal Biostatistician* June 2018-Aug 2018

- Conducted an extensive systematic review of phase III clinical trials that studied pancreatic adenocarcinoma; Summarized all the trial information, the statistical features, and results
- Built a virtual library that contains all the available trials of pancreatic adenocarcinoma from phase II to III; Virtualized hazard ratios across completed studies with multiple presentation talks

Department of Kinesiology and Community Health, University of Illinois **Champaign, IL**
Statistician; Supervisor: Ruopeng An, Ph.D. Jan 2017-May 2017

- Performed data collection on multiple major databases; Performed data management in SAS and R with data cleaning, concatenation, and merging among multiple data sets
- Conducted geographically weighted regression to model the heterogeneity of the association in county level, and virtualized the results via ArcGIS

TEACHING EXPERIENCE

Department of Biostatistics, Johns Hopkins University **Baltimore, MD**
Statistics Tutor and Teaching Assistant Sep 2018-
Supervisor: John McGready, Ph.D., Marie Diener-West, Ph.D. and Leah Jager, Ph.D.

140.623-624 Statistical Methods in Public Health III-IV
140.611-612 Statistical Reasoning in Public Health I-II

- Spearheaded office hour with statistical questions from students: Hypothesis Testing, Linear Regression, Logistic Regression, Poisson Regression, Cox Proportional Hazard Model, Sample Size Determination, Propensity Score Matching, Variable Selection, and Model Selection
- Prepare supplemental lecture notes and explain statistical terminology to students without strong statistical background
- Hold consultation sessions for course projects with statistical advices and writing suggestions

PROJECTS

Data Science

Machine Learning Coursework: Analyzed the NHANES 2003-2004 data with a goal of mortality prediction. Implemented multiple methods that included Random Forest, Boosting, Support Vector Machine, K-Nearest Neighborhood, Linear Discriminant Analysis, and Logistic Regression. The final model was chosen based on the cross-validated test MSE

Data Science Coursework: Applied Generalized Additive Model and Conditional Autoregressive Model to estimate the spatiotemporal effect of PM 2.5 on preventable hospitalization rate in California. Built an R-shiny to present map animations and analytical results

Longitudinal Data

Analysis of Longitudinal Data Coursework: Studied the Children of Immigrants Longitudinal Study (CILS) data to estimate the change of status in feeling discriminated over time among children, and determining features that affect the temporal trends; we also compared the odds of transitioning to and maintaining feelings of discrimination during different life stages.

2018 Qualification Exam Project: Diarrhea risk prediction for a longitudinal study among children in South Asia with a Generalized Linear Mixed Model

Statistical Control and Epidemic Detection

Undergrad Honor Research: Conducted a literature review with a focus on biosurveillance methodology with non-homogeneous observations; generated four extended versions of the Shewhart Algorithm and compared the statistical powers to results from the CUSUM Algorithm, with pre-assigned disease incidence rate ratio.

ACHIEVEMENT

2018 Kocherlakota Award: awarded by the Johns Hopkins Department of Biostatistics for the best performance on the first-year master's qualification exam

2017 Illinois Geometry Laboratory Research Award: awarded by the UIUC Department of Mathematics for exceptional achievement in undergraduate mathematical research

2016 8th SISMID Scholarship: awarded by the University of Washington Department of Biostatistics for outstanding students to attend the summer institute in Statistical Modeling in Infectious Diseases

2016 8th SISMID Certificates: Mathematical Models for Infectious Diseases; Stochastic Epidemic Models and Inference; Statistical Modeling with Novel Data Streams

2016 Society of Actuaries (SOA) Case Study Challenge Certificate of Excellence

2015 AXIS Illinois Case Study Bronze Medalist: awarded by the AXIS Capital at University Research Park

2015 UIUC ISSS Essay Contest Honorable Mention: awarded by International Student and Scholar Services

PUBLICATIONS

An, R.; **Li, X**.; Jiang, N. Geographical Variations in the Environmental Determinants of Physical Inactivity among U.S. Adults. International Journal of Environmental Research and Public Health. 2017; 14(11):1326. doi:10.3390/ijerph14111326

Bronski, J. C., He, Y**., Li, X**., Liu, Y., Sponseller, D. R., & Wolbert, S. (2017). The stability of fixed points for a Kuramoto model with Hebbian interactions. Chaos: An Interdisciplinary Journal of Nonlinear Science,27(5), 053110. doi:10.1063/1.4983524

SOFTWARE

Proficient: R Intermediate: SAS Base, Excel, ArcGIS Beginner: SQL, Access, SAS Macro